comgate

Synchronní replikace v PostgreSQL 9.1

Jakub Ouhrabka

CSPUG setkání

21.6.2011

© ComGate Interactive, s.r.o.

comgate

Proč?

- Durability
- High availibility
- Řešení bez synchronní replikace
  - na aplikační úrovni
  - na úrovni storage (SAN, DRDB…)

© ComGate Interactive, s.r.o.

comgate

Základní vlastnosti synchronní replikace v PG 9.1

- Fyzická replikace (přenos WAL)
- Single master (=primary)
- Multi slave, lze řídit, které slave jsou synchronní
- Řízení sync vs. async na úrovni transakcí
- Slave servery nelze řetězit (PG 9.2)

© ComGate Interactive, s.r.o.

comgate

Možnosti fyzické replikace v PostgreSQL

- Přenos WAL mezi primary a slave
  - File based log shipping
  - Async streaming (9.0)
  - Sync streaming (9.1) - NOVÉ

- Spouštění dotazů na slave
  - Warm standby – nelze spouštět dotazy
  - Hot standby – lze spouštět dotazy

© ComGate Interactive, s.r.o.

comgate

Synchronní COMMIT v PG 9.1

- Požadavek na COMMIT (na primárním serveru)
- Zápis do WAL na primárním serveru
- Přenesení na slave server(y)
- Čekám, dokud nedostanu odpověď
  - async (PG 9.0, default PG 9.1) – nečekám ☺
  - recv WAL
  - sync WAL (PG 9.1)
  - apply WAL
- Vrátím potvrzení COMMITu klientovi

© ComGate Interactive, s.r.o.

comgate

Na co se ne/čeká?

- Na primary serveru se čeká na odpověď slave serveru na
  - COMMIT
  - 2PC PREPARE
  - 2PC COMMIT
- Nečeká se v případě
  - read-only transakce
  - ROLLBACK
  - SAVEPOINT (commit subtransakce)

© ComGate Interactive, s.r.o.

comgate

Co se může stát I

- „Slave je napřed“
  - zápis WAL na primary
  - fsync slave
  - apply slave
  - dotaz na slave, vrací nová data
  - reportování COMMITU primary klientovi

© ComGate Interactive, s.r.o.

comgate

Co se může stát II

- „Slave je pozadu“
  - zápis WAL na primary
  - fsync slave
  - reportování COMMITU primary klientovi
  - dotaz na slave, vrací stará data

© ComGate Interactive, s.r.o.

comgate

Co se může stát III

- Klient neví o provedeném COMMITu I
  - zápis WAL na primary
  - pád primary

- Klient neví o provedeném COMMITu II
  - zápis WAL na primary
  - fsync slave
  - pád primary

© ComGate Interactive, s.r.o.

comgate

Na co si dát pozor I

- Snížení spolehlivosti
  - 2 servery, každý spolehlivost 99,7%
  - Výsledná spolehlivost 99,4%

- Používat minimálně 3 servery

© ComGate Interactive, s.r.o.

comgate

Na co si dát pozor II

- Musím zvládnout situaci, kdy je k dispozici pouze primární server (transakce čekají)

- Vypnout synchronní replikaci, reload konfigurace

© ComGate Interactive, s.r.o.

comgate

Na co si dát pozor III

- Umět se zotavit z pádu primárního serveru
  - Mám několik slave serverů
  - Musím se umět rozhodnout, který budu považovat za primární

- `SELECT pg_last_xlog_receive_location();`
- `-- ne SELECT pg_last_xact_replay_timestamp();`

© ComGate Interactive, s.r.o.

comgate

Na co si dát pozor IV

- Rychlost
- Do každého COMMITu se vkládá navíc network round-trip

- Vybírat si transakce, kde chci použít synchronní replikaci a kde ne

© ComGate Interactive, s.r.o.

comgate

Konfigurace synchronní replikace

- Primary - postgresql.conf
  - `synchronous_standby_names = * nebo jména`
  - `synchronous_commit = on/local/off` (lze i pomocí SET)
  - `wal_level = archive/hot_standby (minimal)`
  - `archive_mode = on`
  - `max_wal_senders`

- Primary - pg_hba.conf
  - povolit připojení k databázi `replication`

- Slave - recovery.conf
  - `standby_mode = on`
  - `primary_conninfo, application_name`

- Slave - postgresql.conf
  - `hot_standby = on/off`

© ComGate Interactive, s.r.o.

comgate

Jak zjistím stav?

- SELECT * FROM pg_stat_replication
  - application_name
  - state STREAMING/CATCHUP
  - sync_priority
  - sync_state ASYNC/SYNC/POTENTIAL

  - sent_location
  - replay_location

- `wal_receiver_status_interval`

© ComGate, a.s.

comgate

pg_basebackup

```
> pg_basebackup -D /tmp/newcluster -U replication –v

NOTICE: pg_stop_backup complete, all required WAL
segments have been archived

pg_basebackup: base backup completed
```

© ComGate Interactive, s.r.o.

comgate

recovery.conf

```
standby_mode = on

primary_conninfo = 'host=localhost port=59121
   user=replication password=replication
   application_name=newcluster’
```

© ComGate Interactive, s.r.o.

comgate

Spuštění slave clusteru

```
pg_ctl -D /tmp/newcluster start

...

LOG: streaming replication successfully connected to
  primary
```

© ComGate Interactive, s.r.o.

comgate

Nástroje + zdroje

- Nástroje
  - http://projects.2ndquadrant.com/repmgr
- Zdroje
  - http://www.postgresql.org/docs/9.1/static/
  - http://www.depesz.com/index.php/2011/03/18/waiting-for-9-1-synchronous-replication/
  - http://tech.myemma.com/replication-synchronized/

© ComGate Interactive, s.r.o.